4. DATA VIEW

The Data View characterizes the data required to support the ICS functions as described in the Functional View Section of this document. The main data object in ICS is the *Collection.* Therefore, the main focus of this section is a description of the conceptual and logical nature of the *Collection.* To begin, Section 4.1 provides a motivation from an EO user's point of view for *Collections.* This section is then followed by a description of the *Collections* Model and a brief description of using *collections* with earth observation data. Section 4.4 addresses ction Concept Details while Section 4.5 describes how the users will interact with the model. This would then conclude the conceptual discussion of the *collection.* Section 4.6 primarily addresses the logical nature of the *Collection* by providing an overall description of the data components of ICS. This description includes an illustration of the ICS data components coupled with a brief textual narration. Section 4.4 briefly describes the document data.

4.1 Motivation for Collections

ICS Compatability: Explanatory

Users of Earth Observation data will be interested in subsets of data from various datasets and various locations. A user is typically interested in an assembly of data which best fits the topics which the user needs to investigate. This assembly requires the data to be organized in a certain way. It is this notion - that the users need for data does not uniformly overlap the organization - that provides the motivation for the collection concept.

The following are some examples of how EO Data consumers will want to access EO data:

- It is assumed that in general the user community will often desire access to specific groupings of ICS holdings. For example the user may request information from a single instrument, or for a specific wind-speed and direction data from AMSR, or of trace gas distribution data from MIPAS;
- Because agencies have responsibility for holding the data from certain sources, the physical location of data will be assumed to be geographically distributed around the world. Collections can be formed which can be associated with products from multiple geographic locations;
- It is assumed that the community will also require access which spans instruments. For example, a request for all sea surface temperature data, across MODIS, AMSR, and so on;
- It is assumed that inter-disciplinary work will call for access across science disciplines and instruments, relative to some specific time, space, or event coincident coverage.

Therefore, to address these somewhat divergent but related needs the concept of a *collection* was developed. This concept embraces the following fundamental concepts.

- The *collection* is an item
- The *collection* has members
- The members can be other items

These concepts allow the data to be organized in a manner which will facilitate the access needs of the user.

The following sections explore these concepts in further detail by introducing a model of the collections concept, describing the collection membership, and providing examples of user interactions with the collections.

4.2 Collections Model Overview

ICS Compatability: Explanatory

The collections concept permits a wide degree of flexibility in organizing EO data. A *collection* is described in ICS as a collection descriptor, i.e., a set of elements which describe the collection. The *collection* may have members which are reflected as included item descriptor identifiers. These identifiers serve to identify the associated *products,collections,* and *task packages,* through the *Included Product Descriptors,* , *Included Collection Descriptors,* or *Included Task Packages* respectively.

A *collection* supports the notion of associated data or membership either due to their physical location (i.e. database storage), or along the lines of subject themes, that are deemed useful organizational principles. The science data provider coupled with a *Retrieval Manager Administrator* should decide the appropriate organizational strategy. Certain rules will be specified in the ICS Collection Manual to provide the necessary guidance to assist this team in making the appropriate decision.

Some examples for arranging data follows. Traditionally EO data was arranged according to EO product types (Raw, Level 0, etc.). Within ICS *Collections*, data can be arranged according to agency, discipline, spectral range, instrument, processing level, geographical area, etc. The user can then direct a search in a more focused manner, by including or excluding particular collections. As a *collection* can contain either included *product descriptor* ID's, *included collection descriptor* ID's or included *persistent result set* or included *persistent query* , it is possible to associate a number of collections under a single theme as the data provider or user finds applicable. It is also possible to associate an existing set of *product descriptor* ID's, which are already associated to an existing *collection*, to a new *collection* by just referring to the existing *collection descriptor* ID within the definition of the new *collection descriptor*. Because Release B of CIP does not support mixed collections each *collection* may only include descriptors of a specific type; i.e. *Product, Collection , persistent query* or *persistent result set..*

Figure 4-1 illustrates this concept followed by Figure 4-2 which further illustrates the notion of a *collection* and its' relationship (association) with other collections.

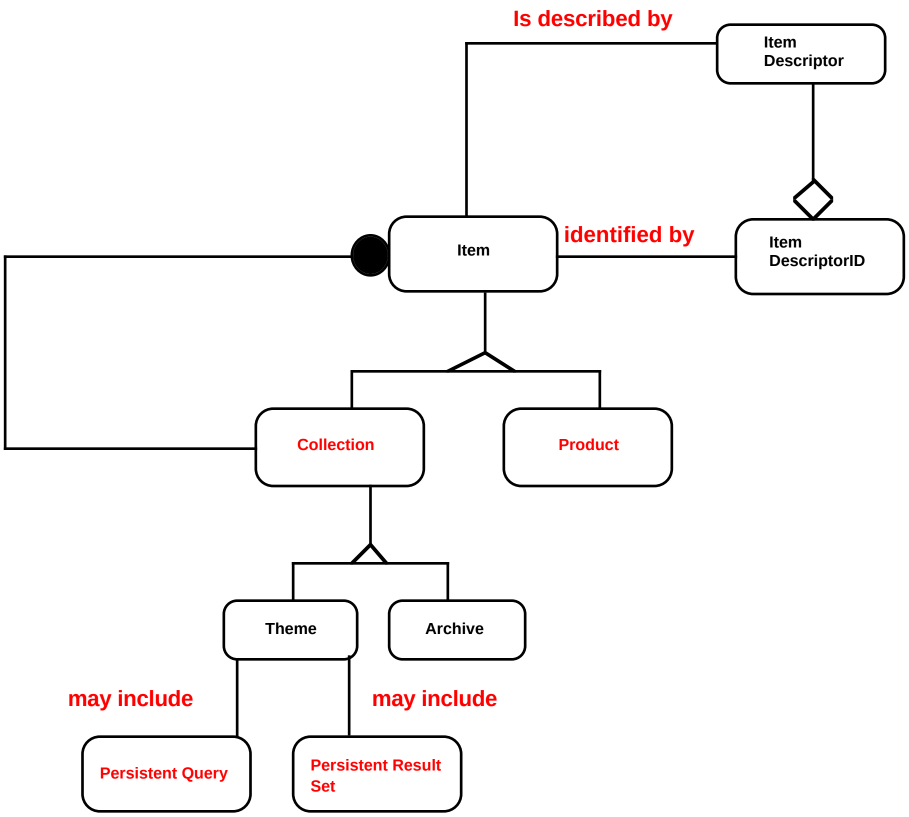


Figure 4-1. Single Instance of Collection in Figure 4-2.

The above model is an attempt to graphically represent the concept of a *collection* which is illustrated in Figure 4-2. First an item is equivalent to a collection in Figure 4-2 (i.e. collection 2.2). Each item is described by an item descriptor which logically consists of a group of elements which describe the overall nature of the item, such as, name, temporal and spatial extents. The items can be either a collection or product; the collection item can be either a theme or archive. Each item is identified by a unique item descriptor. Collections may contain zero or more items. These items can be other collections or products. The remaining sections will explore each of the above elements of the collection.

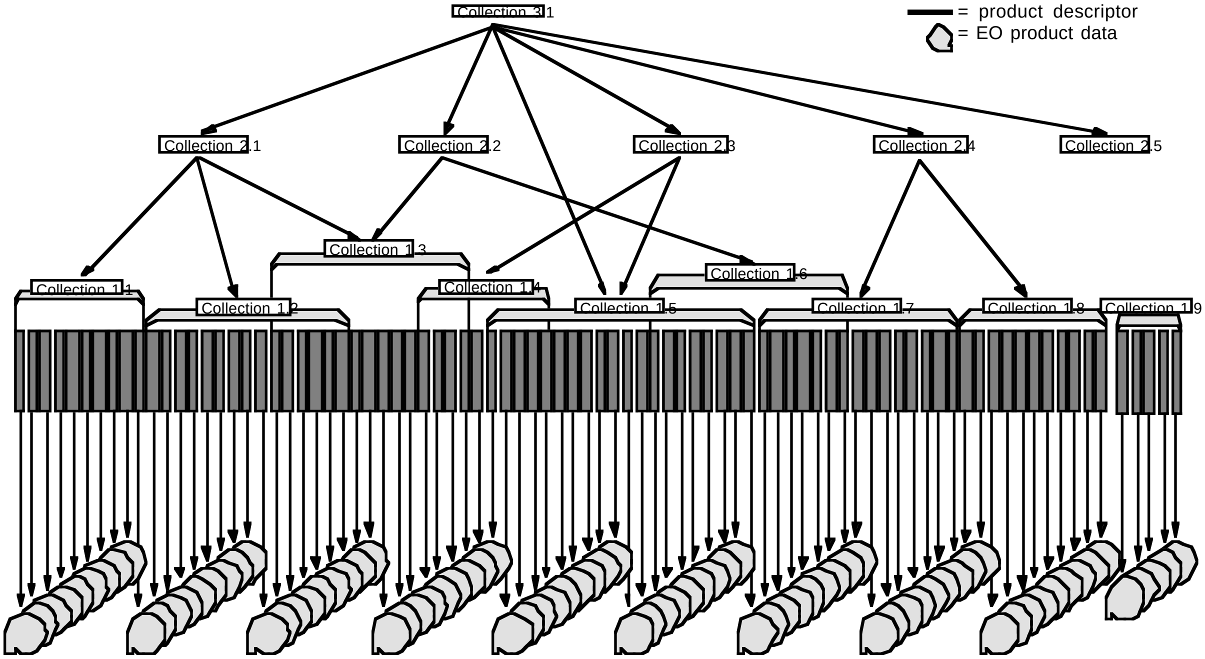


Figure 4-2. The "Collection Structure" Concept

Figure 4-2 expresses the logical relationship between multiple *collections*. As can be seen a web of interesting relationships can be formulated when using the collection concept. The collections in Figure 4-2 do not represent the naming of collections in an actual implementation. As can be seen the collections can overlap each other and can be associated with more than one collection. The central concept is that there are terminal collections, i.e., Collection 1.2 and non-terminal, i.e., Collection 2.2. The terminal collections in this illustration describe the information contained in the inventories and are defined as Archive Collections in ICS. The Archive Collection Included Item Descriptor Id's will reflect only the Included Product Descriptor ID's. Above the terminal level collections, there are non-terminal collections. These collections are typically constructed around a theme, or event, however not limited to these aggregations, which expresses the primary tenet of the associated collections. In ICS these are the Theme Collections. The included collections may be any existing collection.

There is no restriction on the number of included item descriptors within collection levels, therefore a non-terminal collection could group together terminal collections and other non-terminal collections. Also, a terminal collection could exist without a relationship to a higher collection (i.e. collection 1.9), or a non-terminal collection could exist with no relationship to lower collections, in other words a collection without members (i.e. collection 2.5). These conditions could occur in particular circumstances, for instance when a catalogue site has only one single collection, i.e. Collection 1.9, or when the collection is under construction, or is a Theme Collection, i.e., Collection 2.5. The specific details for constructing and maintaining the collections are described in the ICS Collections Manual.

An integral part of the collection structure is the Root Node. The overall purpose of the root node is to allow a *Search Request*, which was formulated at the origin by a user, to ultimately determine the EO *collection* that match the criteria specified in the query attributes of the *Search Request.* The root node serves as the logical mechanism which supports this service. The criteria specified in the query are evaluated against the contents of the Root Node and all of the Root Nodes' included collections to determine if the *collections* in the targeted *collection* tree satisfies the *Search Request*. The data objects model for the root node contains the *collection* descriptor object and attributes.

Specific rules and the required procedures for developing and maintaining a root node are identified and discussed in the ICS Collections Manual.

4.3 Example Of Using Collections With EO Data

ICS Compatability: Explanatory

Several examples of constructing collections based on the users needs are reflected in Figure 4-3. In this illustration a Collection has been established to include MODIS data from several missions. Also a Collection has been created based on the theme Sea Surface Temperature (SST), which associates existing SST collections from AMSR, AATSR, and MODIS. The intent of the subset caption is to show that only some of the item descriptors are included in the MODIS collection or AMSR on ADEOS II, or AATSR on ENVISAT, or SAR on Radarsat. Lastly, a Collection is defined based on an Andes Event. As illustrated, this collection includes collections which otherwise may not be related.

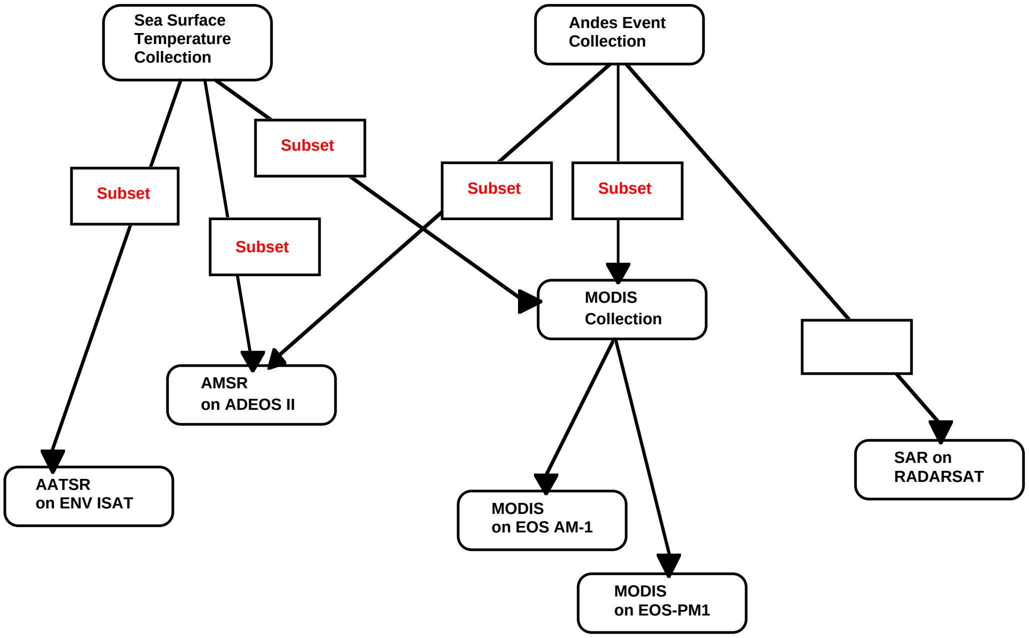


Figure 4-3. Creating Collections for Users

4.4 Collection Concept Details

In order to fully understand the power of the Collection Concept within ICS the following detailed information is provided in this section. The intent is to describe the collection descriptor classifications (categories), to provide some insight into collection membership, to discuss and define characteristics which are unique to this concept, and lastly, to describe how a user can interact with collections.

4.4.1 Collection Descriptor Classification

ICS Compatability: Explanatory

The flexibility of the collections concept requires an accurate classification scheme to allow the users of ICS to determine the characteristics of certain collections they are creating or viewing. In support of this requirement two major classification elements are used to classify *collections* in the ICS. The following describes each of these elements.

Type

- *Archive* - An *archive collection* reflects a single underlying physical archive that is "owned" by a single *Retrieval Manager*
- *Theme* - A *theme collection* reflects a common concept or theme and may include Provider Archive or Other Theme Collection members that are "owned" by several different *Retrieval Managers*.

Examples of collections which can be used to group data together which have a similar semantic theme can be seen in Figure 4-4 . The classification of those *collections* include:

1. *Collections* of SENSOR data by mission (e.g. MODIS from AM-1, SAR from RADARSAT, etc.) which reflects the organization of the underlying physical archives. *archive* would be the correct element classification.

2. A *collection* has been established to include MODIS data from several missions. A collection has been created based on the theme of Sea Surface Temperature (SST) including SST products from AMSR, AATSR and MODIS. This collection would be classified as a *Theme* collection.

The first type of *collection, Archive*, is likely to be created by data providers to organize their archives and facilitate access to the *product* descriptors (i.e. analogous to an inventory containing inventory entries). The second type of *collection, Theme*, may be established by data providers or Users who want to organize some of their data into groupings which differ from their *archive collections* (i.e. from the baseline inventory), for the convenience of their users. The following provides a more detailed explanation of the provider archive and theme collections.

Archive collections. An *archive collection* could be a dynamic collection where new item descriptors are automatically added to the collection by the catalogue system as new archive entries are created, i.e. for data from an existing, operational EO satellite. Alternatively, an *archive collection* could be a static fixed collection of historical data that is no longer being added to as the source satellite is no longer operational or the campaign is no longer active. An *archive collection* includes only item descriptors that are local.

Theme collections. This *collection* could be based on a covered geographical area, a scientific discipline supported by the data, or an instrument type, etc. The difference between *theme* and *archive collections* is that *archive collections* only contain homogeneous information, whil*e Theme Collections* may have item descriptors whose contents may vary depending on the nature of the collection. *Theme Collections* are non terminal. Theme Collections may also be created to include unregistered collections that have been created and preserved in the Extended Services Database. All of these options are dependent on how the Data Administrator decides to organize the data.

Note that these classification definitions are not mandatory for the CIP to operate, however to ensure data interoperability the *collections* described above must share identical semantical interpretation across ICS. They must also be under the control of the ISA which is one step towards ensuring interoperability. The *Retrieval Manager* makes use of *collections* for the routing of distributed queries. Although a *collection descriptor* data element is identified and defined to capture a collection classification, the primary intent of the categorization is to aid the ICS user in

understanding the quality guidelines imposed on the *collection* data and metadata. Standardization of *collection* definitions is provided as part of the ICS design. Guidelines for the creation and maintenance of these collections in addition to further explanation of the characteristics of these collection classifications can be found in the ICS Collections Manual.
The two classifications of collections are illustrated in Figure 4-4.

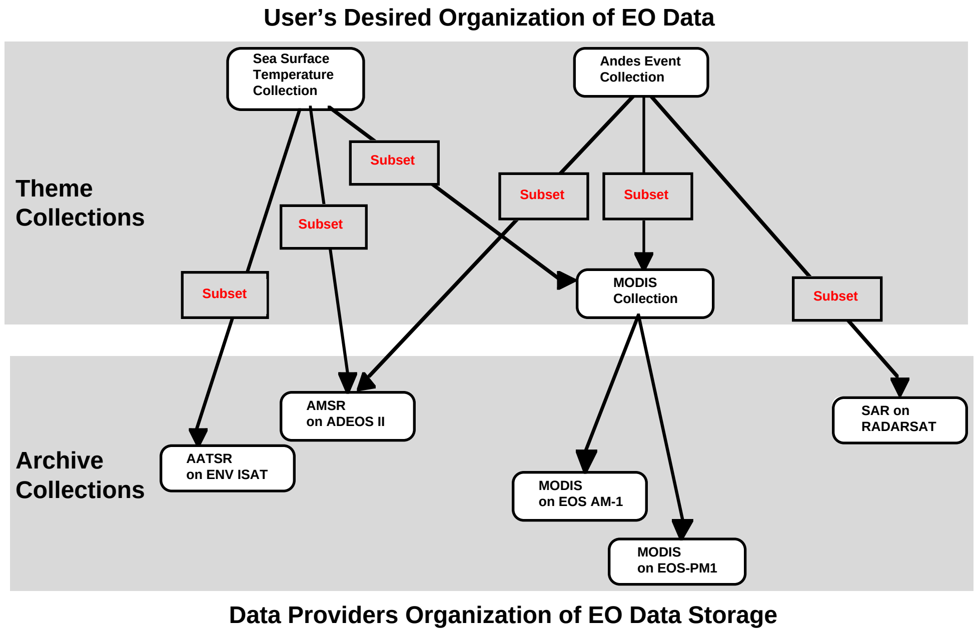


Figure 4-4. Theme and Archive Collections

4.4.2 Collection Members

ICS Compatability: Explanatory

Collection Membership in ICS is defined as including remote or local item descriptor ID's in the *collection descriptor* which defines the *collection*. A detailed process for identifying and including members will be described in the ICS Collections Manual. However, there are several fundamental rules which serve to guide this process.

1. A *product* descriptor identifier may be included in many Theme Collections however as a minimum it must be contained in one and only one *archive collection.*
2. *Archive collections* must include one or more *product* descriptors
3. *Theme collections* may include *collection* descriptors which may be an *archive* or *theme collection* descriptor.

4.4.3 Collection Characteristics

ICS Compatability: Explanatory

This section explores in more detail, the higher level concepts that were introduced in the previous sections. This is achieved by identifying and defining detailed *collection* characteristics to include the following:

- Commonality
- Evolution over time
- Terminal Collections
- Identifier
- Uniqueness
- Remote collections
- Related collections
- Related Guide Documents
- Local Attributes
- *Registered* Collections
- *Unregistered* Collections
- Included Task Packages

Commonality:

By definition, a *collection* is a grouping of items that have something in common. A *collection* may have members that have many or fully common attributes (*archive collection*), or a *collection* may have members that have a common semantic theme, though only a small subset of common attributes (*theme collection*).

Further CIP specifies a list of standard attributes that can be searched. Some of these standard attributes are mandatory for all *collection* members (different mandatory sets for different descriptor types), while some are optional (although commonly understood). Finally, some attributes can be locally defined.

Evolution over time:

Static members do not change over time. This can be a result of static underlying *collections* or the mechanism used to create the *theme collection* such as a Volcanic Eruption Theme Collection whose members will more than likely remain static over time.

Dynamic members may change over time based on changes in the included *collections*. It is envisioned that the majority of *collections* will not be static, but will evolve as ICS is used. This will occur in response to the way in which the user community wants to view relationships between the various data held by ICS, and the ways the CEOS Agencies wish to respond to those desires. Dynamic membership will require close supervision by the *ISC Site Administrator* to ensure that the *collection* descriptor information is current.

Terminal Collection

As *collections* can contain pointers to other collections, reflected in the included Item Descriptor ID's, there exists the concept of a 'collection structure' (see Figure 4-2), the leaves of the branches being product descriptors. The *collections* that include only *product* descriptors are termed 'terminal collections, however, Theme Collections which include Remote Collections may act as terminal collections at the RM that hosts the collection structure.

Identifier (In the following text, *item descriptor* includes *product*, or *collection descriptor):*

Each member of a *collection* (i.e. included item descriptor of any type) must have an identifier unique within all the *collections* in the *retrieval manager's* collection tree. This unique identifier can be seen as the name of the item descriptor. In particular, it will be a single unique identifier which includes the *Retrieval Manager* identifier, the *collection* identifier and the *collection* member identifier(included item descriptor ID's). For a *collection* descriptor, the *collection* identifier refers to a collection of any classification (*archive,or theme ,*), whereas for *product* descriptors, the identifier must refer to an *archive collection*. Specific format details can be found in the CIP Specification[R3].

Uniqueness:

A *collection* member (included item descriptor ID) may be a member of more than one *collection*, as illustrated in Figure 4-4, Provider Archive Collection AMSR on ADEOS II is a member of the Sea Surface Temperature Collection as well as the Andes Event Collection . However, duplicate members (included item descriptor ID's) must not be visible within a single *collection* . For example, Provider Archive Collection AMSR on ADEOSII could not appear twice as an included item descriptor ID in a collection that contained both the Sea Surface Temperature Collection and the Andes Event Collection . This property is known as elimination of duplicates to achieve uniqueness.

In the case of a *collection* which is a child of two or more included *collections*, any operation such as search,which traverses the *collection* tree from the top level collection will end up repeatedly visiting the child *collection*. The unique collection identifier provides a means of preventing repeated operations on the same *collection* . This is achieved by noting which tree nodes (unique identifiers) had been visited and then restricting access to those nodes (unique identifiers) for the same search. This functionality will be performed by the *Retrieval Manager*.

Remote Collections:

Remote collections are members of a collection hierarchy whose information is stored or maintained at a CIP site other than the one in which it is included as an included remote item.

Normally, a *collection* structure would be held in one place (say, as a database on a computer). A logical *collection* tree is where one or more of the members are held elsewhere - the complete *collection* tree thus spans multiple sites. The *collection* that is referencing(included item descriptor or included task package) a *collection* at the remote location is termed a 'remote member'.

Remote members do not have to maintain information about which members refer to them; remote members are indistinguishable from local *collection* members from the user's point of view. This concept is supported by the consistent use of URLs to identify *collections,* in the same manner as the complete WWW is seen by the user as a single database. A *Retrieval Manager* 'owns' those *collections* for which it stores and maintains the attributes, and only stores the pointers(Included Item Descriptor ID's/included task packages) to remote members, not their attributes and values.

No attribute or value of an attribute for a remote member, or the pointer (included item descriptor/task package name) to the remote member, can be guaranteed. The *Retrieval Manager* where the remote member is stored may not be available; the remote member may have changed its data structure (adding, changing or deleting attributes), or the remote member may have been deleted from the remote *Retrieval Manager*.

Related collections:

Collections may be related to one another without the need for a "parent-child" or the "include" construct. The relation may be through content or purpose, for example, and allows the spanning of one *collection* tree to another. A *collection* descriptor will contain a list (possibly empty) of related *collections* as part of its content.

Related Guides

It will be possible to relate Guide Documents to collections within the collection descriptor. This can be achieved by indicating the associated Guide URL in the Related Item ID data element of the collection descriptor.

Local Attributes

Local Attributes are collection specific characteristics about a collection. The existance of local attributes may be specified within a collection descriptor by setting the Local Use Attribute Flag Element. A flag of 0 indicates that the collection does not contain local attributes, 1 indicates that the local attributes are described within the collection descriptor whose values are captured in the Product Descriptor; 2 indicates that the local attributes have been described in the Explain and the values for the attributes are captured in the Product Descriptor. Local Attributes Using the Collections Database and Local Attributes Using Explain are addressed in Sections 4.6.1.2 and 4.6.2.2 respectively. Specific details regarding the creation/maintenance of local attributes are contained in the ICS Collection Manual.

Registered Collection -

Registered Collections contain item descriptors, implies persistence (i.e.) long lived and has a collection descriptor in the Collections Database. May be accessed via the full range of CIP access services including *discovery, navigation, location and searching.*

- *Archive collections*: : This type of collection is likely to be created by data providers to organize their archives and facilitate access to the product descriptors (i.e. analogous to an inventory containing inventory entries. May be static or dynamic
- *Theme collections*: This type of collection may be set up by data providers or users who want to organize some of their data into groupings which differ from their provider archive collections (i.e. from the baseline inventory), for the convenience of their users, for example, based on the geographical area covered, the scientific discipline supported by the data, the instrument type, etc. May be static or dynamic.

Unregistered Collections

Unregistered Collections (also known as result sets) - contains item descriptors, does not imply persistance and does not have a collection descriptor in the collections database. May only be accessed via the locate service or by the creator of the collection.

Persistent result sets - implies persistance, May be static or dynamic. May be accessed using *Locate* service.

Temporary result sets- implies no persistance (i.e. only available in the z-association in which it was created). Is static. Can only be accessed by creator in the same z-association in which it was created.

Included Task Packages

Task Packages may be included as members of a collection. The purpose for including Task packages would be to establish a relationship between the collection and the extended services result sets which may be a persistent query or an accumulation of result set item descriptors. Reference Section 4.6.4 for a description of how included task packages would be used by the Retrieval Manager.

4.5 User Interactions with Collections

ICS Compatability: Explanatory

This section describes multiple methods a user will have available using the CIP within the ICS domain to discover items in *collections*, e.g. *products, guide, browse,* etc. Four methods for using the *collection* structure are described:

- Collection Discovery
- Collection Navigation

- Collection Searching.
- Collection Locating.

Each of these methods differ based on what the user needs to know before using the method and what element of the *collection* descriptor is used by the method. An example is that for *collection* discovery the user needs to know very little of the ICS, compared with collection locating in which the user must have a specific *collection* ID to use the method.

These methods are depicted in Figure 4-5 which illustrates each of the *collection* usage methods identified above relative to a *collection* node. Each of the *collection* usage methods are described in the remainder of this section.

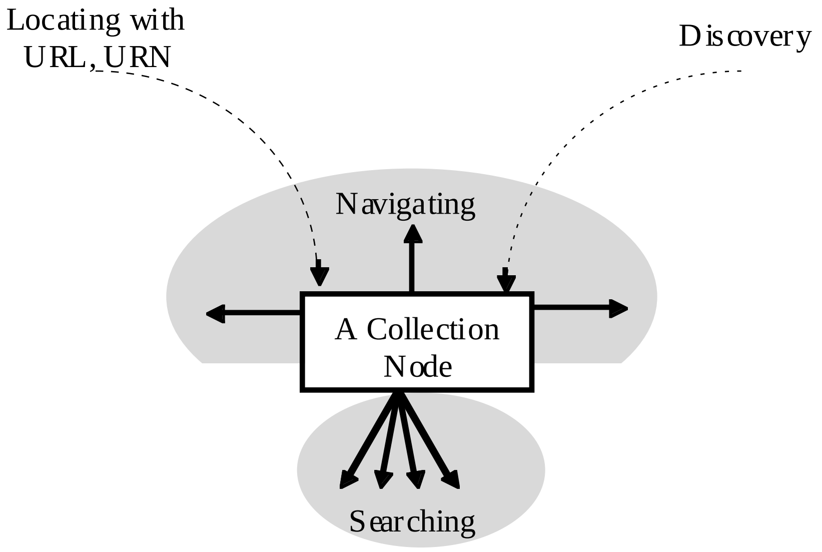


Figure 4-5. Collection Usage Methods

Collection Discovery:

Collection discovery allows a user of the ICS, having no prior knowledge of the *collection* structure, to be assured that all *collections* of interest to the user are found. It is assumed that the user will know of at least one *Retrieval Manager*. This method provides a user with an exhaustive method to determine all *collections* which may be of interest. The discovery of *collections* makes no assumptions about the user knowledge of any existing *collections* or other *Retrieval Managers*.

For CIP-B, *collection* discovery will be accomplished by the relatively simple method of searching a global *collection* which is discussed in Section 3. It is anticipated that information discovery methods currently being researched in WWW community may have value for accomplishing *collection* discovery in later releases.

Collection Navigation:

Collection navigation allows the user to examine *collections* related to a specified *collection* and switch context to the related *collection*. This is achieved by allowing the user to navigate through a collection structure resident on the user's client. When the user reaches the edge of the collection structure a search request will need to be made to the RM to retrieve the collection of interest. The related item descriptor is the data design construct which supports this functionality. The Collection Navigation service will apply only to the Archive collections and Theme Collections.

Collection Searching:

Collection searching is the core function of the CIP. With this method the user knows of a *collection* of interest as well as the *Retrieval Manager* where the *collection* descriptor for the *collection* resides. The user establishes a session with the *Retrieval Manager* and targets a search at the *collection* The search is then propagated down through the included item descriptors of the *collection* and any included *collections*.

Collection Locating:

Collection locating assumes that the user has a *collection* URL in their possession and wishes to find the *collection*, examine the *collection* elements, and perhaps target a search at the *collection*. Having a URL, the user can examine the URL and determine information about the *collections* location, the user can submit the URL to a *Retrieval Manager* which can locate the *collection* or determine that the *collection* no longer exists.

The following table summarizes the salient concepts which characterize the various collections that have been presented throughout Section 4.2.

The **Owner** is the individual(s) responsible for the management and ownership of the collection. **Ref in Coll Structure** implies that the collection can be included in an existing collection. **CIP Metadata** refers to the existence of data about the collection in the Collections Database. **Lifespan** is the anticipated longevity of the collection. Please note this is the anticipated longevity. Collections may be deleted under certain circumstances as identified in the CM. **Services** are the methods that may be used to access the collections and lastly **Attributes Homogeneity** refers to the likelihood that the characteristics which describe the collection will be similar.

Table 4-1. Collection Concepts Summary

	Owner	Ref in Collection Structure	CIP Metadata	Lifespan	Services Capable of Discovery	Attributes Homogeneity	Static or Dynamic
Registered Collections							
Archive Collection	Admin.	Yes	Yes	Long Term	Discovery Navigate Search Locate	High	Based on specific collection
Theme Collection	Admin.	Yes	Yes	Long Term	Discovery Navigate Search Locate	Medium	Based on specific collection
Unregistered Collections (i.e.result sets)							
Persistent Result Set	Admin. / User	Yes	No	Medium Term	Locate	Low	Based on specific implementation or collection
Temporary Result Set	User	No	No	single Z association	None	Low	Static

4.6 ICS Data Framework

This section of the Data View will present the logical data components of the ICS. This will be achieved by describing and defining each data component within the overall Data Framework. This framework as illustrated in Figure 4-6 represents the logical data objects which will be managed and maintained in ICS. The emphasis, in the illustration, and this section, is on the data held in the *Retrieval Manager* which includes the:

- *Collections Database (CDB)* (Section 4.6.1),
- *Explain Database* (Section 4.6.2),
- *Extended Services Database* (Section 4.6.3),
- *Result Set Database (*Section 4.6.4).
- Session, Error and User Data (Section 4.6.5-4.6.7).

The data in the IGP Domain are defined in the Guide Design and Protocol Specification Document. However, Section 4.7 of this document provides an overview of the ICS Document Data.

The data required by elements other than the *Retrieval Manager*, for example external interfaces, are recognized in Section 4.7. Each data component will be described in the sections indicated above. These descriptions will include an overview of the contents of the database, an object model, which will illustrate the data objects, attributes and relationships for each of these data repositories, and lastly any unique design characteristics surrounding these databases. However, the CIP schema described in the CIP Specification [R3] is the definitive source for all schema definitions. Additionally, the valid domain of values associated with each attribute is identified and described in the ICS Valids Document [R4].

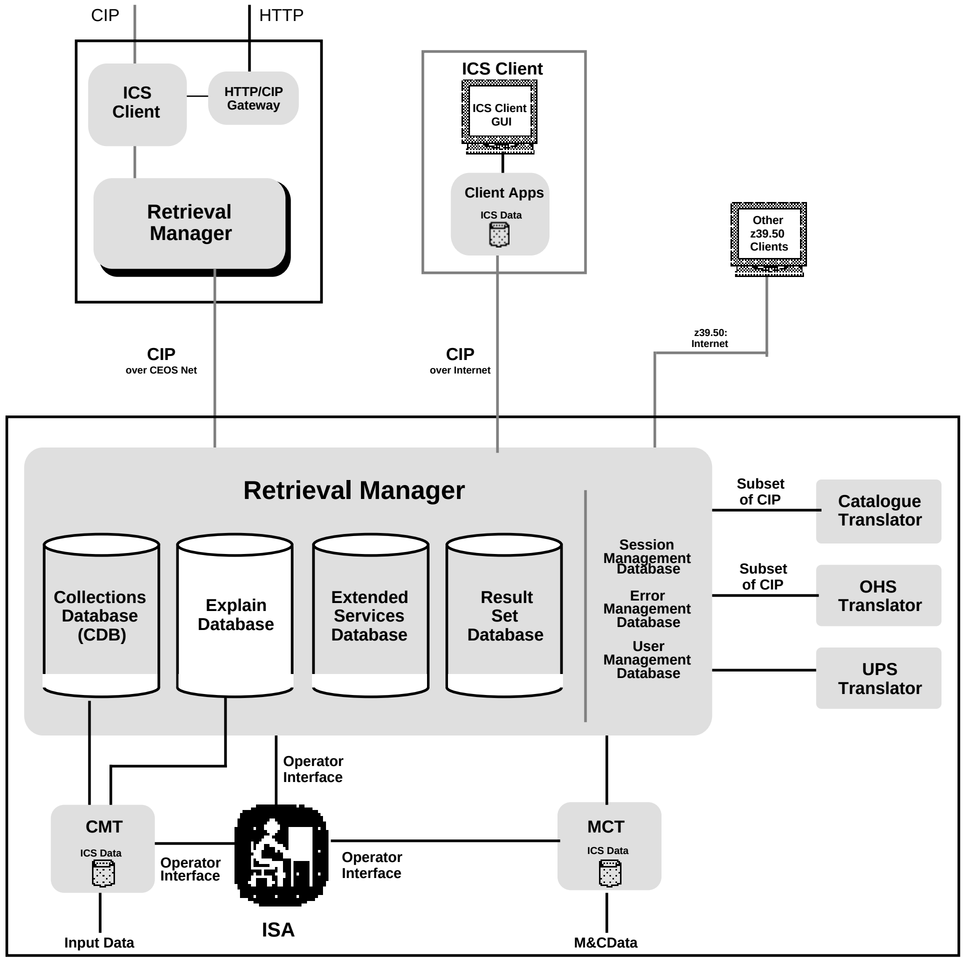


Figure 4-6. ICS CIP Data Framework

4.6.1 Collection Database(CDB)

The *collection database* contains the instances of the science metadata which are referenced from the model illustrated in Figure 4-1. The metadata coupled with the organization of the metadata within the *collections database* is used to assist the user in discovering, searching, and ordering Earth Observation information. The intent of the *collections database* is to satisfy these objectives within a framework which will allow maximum flexibility while minimizing change to the overall structure. This is achieved through the item descriptor which is a generalization of the specific classes of data which are supported under the *collection database* umbrella. The item descriptors supported within CIP's *collection database* are the *collection,and product*, Descriptors. The following definitions provide the context for each of these descriptors.

Collection Descriptors	Collection metadata serves to characterize the underlying product information/included collections. These characteristics consist of spatial, temporal, source, quality, document references, and keyword informational elements.
Product Descriptors	Product metadata characterizes the individual products contained in the archives. Each product can be described in terms of it's spatial, temporal, source and quality attributes.

4.6.1.1 Collection and Product Attributes

ICS Compatability: Explanatory

For any of the above descriptors, a standard set of attributes are required (see CIP Specification - Release B [R3]). Note that not all catalogue systems will support all attributes, but the minimum set of attributes as specified in the CIP Specification should be supported. In the case of *product* descriptors (the attributes of which are based upon currently existing inventory entry definitions), there may be attributes that are local to a particular catalogue system. The CIP will permit the transfer and searching of these attributes without needing to 'understand' their real world meanings.

An example of how the attributes can be used is presented in the tables below. (Note that the attributes represented in these examples are for illustrative purposes only. These examples do not include the full list of mandatory attributes. The complete result schema is defined in the CIP Specification Release B, Appendix C [R3], which indicates those attributes that are mandatory or optional.

The list of *product* descriptor attributes that might describe an AVHRR data product, and the corresponding values for a single instance of that product, might be:

Product Descriptor Attribute	**Example Value**
SensorName	AVHRR
EndDate	1994-07-10T17:20:43.123456Z
MissionId	NOAA-09
SpatialCoverage	-32.27, -79.39, 62.55, 65.32
ArchivingCentreId	OBERPFAFFENHOFEN
ItemDescriptorId	PID_103.47c6dec1

Finally, the EO collection descriptor for a *theme collection* may be part of a *collection* that contains all AVHRR data covering the European continent, a list of the *collection* descriptor attributes and their values that might describe such a collection could be:

Collection Descriptor Attribute	**Example Value**
ItemDescriptorId	cip://ciprm.esrin.esa.it/CID_121
ItemDescriptorName	AVHRR
CreationDate	1994-05-01T00:00.0
CollectionHierarchyPosition	NON-TERMINAL
CollectionCategory	Theme Collection
Purpose	This collection is a grouping together of all known AVHRR data that is available via the ESRIN and DLR EO data centres.
RevisionDate	1995-01-15T00:00.0
VersionId	2.3
SpatialCoverage	-10, -70, 35, 70
Progress	IN WORK
BrowseId	http://gds.esrin.esa.it/avhrr/collections/europe.jpg
IncludedCollectionDescriptors	cip://ciprm.esrin.esa.it/CID_226 cip://ciprm.dfd.dlr.de/CID_125

The list of attributes for each of the above item descriptor objects (*product* descriptor, and *collection* descriptor) will be different, as the information carried by these objects is inherently different. It is also unrealistic to assume that (apart from any mandatory subset) the list of attributes for any one object within one agency will be exactly the same within another agency; this is especially true for the *product* descriptor objects, where existing archives and inventory entries are very different.

The CIP Specification - Release B specifies a set of attributes for describing *collections*. Not all servers need to support all *collection* attributes, however a mandatory set must be specified (see Appendix C of [R3]). For *product* descriptors there are several mandatory attributes that shall ensure basic interoperability between *product* descriptors (see Appendix C of [R3]).

4.6.1.2 Local Attribute Definition Using the CDB

ICS Compatability: MAA

Each CIP Archive Collection, can be further characterized beyond the standard set of elements specified in the CIP Specification [R3]. This is achieved through the identification and description of local attributes. This capability will allow the developer of the collection to identify and record, either in the Explain Database and Collection Database, or only in the Collection Database, additional characteristics about a collection such as AverageBlack Body Temperatures, to give an example.

The process and overall procedure for capturing and recording this information is described in the Collections Manual. The data model for the Collection Database, Collections Module and further the local attributes are specified in Figure 4-7.1. From this figure one can determine the information that is necessary to describe the collection specific data that can not be described using the standard set of attributes. The key ISA decision surrounding the implementation of the local attribute concept is whether to record the local attributes characteristics in the Collection Database or in the Explain Database. These characteristics are the explanations of the local attributes such as the definition of the local attribute, the meaning, the name etc. The local use attribute flag element which is contained in the collection descriptor expresses the ISA decision and is used by the search and present services to determine the target (Collections Database or Explain) for the request.

4.6.1.3 CDB Support for CIP / GEO Interoperability

ICS Compatability: Explanatory

The collections database schema currently contains the necessary elements and use attributes to support data interoperability between the GEO and CIP. This was achieved by aligning the data between the two applications, paying careful attention to the semantical and structural differences and identifying a set of common characteristics which may be applied to both applications. Due to this alignment users of either CIP or GEO can formulate queries which may target each other and expect semantically similiar results. The specifics surrounding the data and functionality required to ensure this interoperability can be found in the CIP/GEO Specification [R].

4.6.1.4 ICS Valid Values

ICS Compatability: Mandatory(See R4)

The ICS has an associated list of valid values which are specified allowable terms for a select group of attributes. These valid values serve the following important purposes.

- Establishes a clear definition of the use attributes by providing a list of examples (valid values) that maybe used in populating the attributes.
- Maintains the integrity of the data for searchable attributes.
- Ensures data interoperability across sites where a variety of data value solutions could be applied to a given use attribute.

The ICS Valids Document [R4] contains the list of attributes and associated controlled list of valid values. This document also specifies the ICS procedures for maintaining the valids.

4.6.1.5 Collections Database(CDB) Data Model

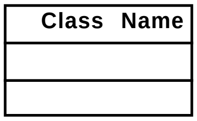
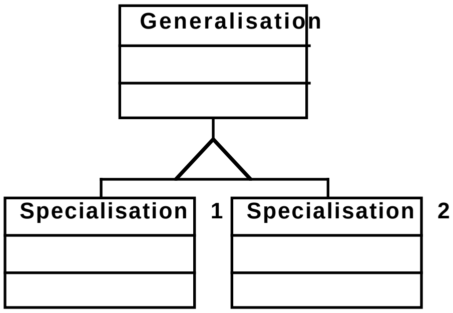
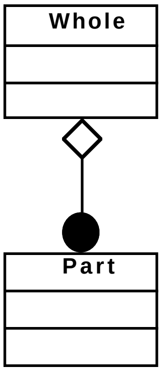
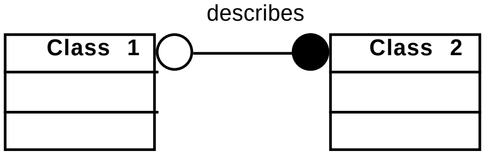
ICS Compatability: Mandatory(See R3)

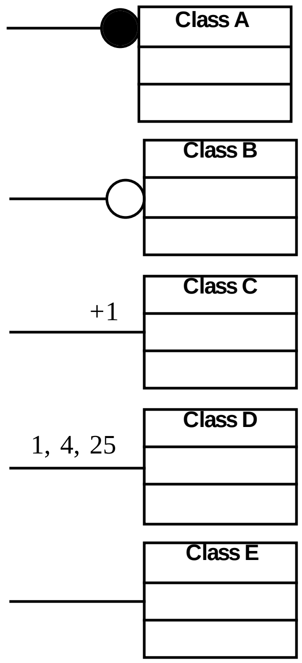
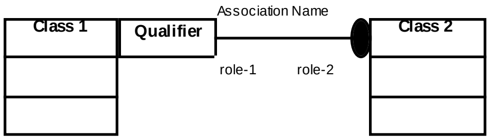
The Collection Database Data Object Model, which is illustrated in Figure 4-7.1 through 4-7.5, expresses the EO data objects, attributes, cardinality, and relationships between the objects. There are five modules which collectively represent the ICS CDB. The modules allow a partitioning of the entire CDB into manageable subsets. These models attempt to illustrate the relationship between the collections and other CIP objects such as the Explain Database and the Extended Services Database.

Description of the CIP objects and attributes are recorded in Appendix “B” and “C” of the CIP Specification - Release B [R3]. The ARS’s which is the basis for these models is also contained in the Rel B Specification, Section C.3.[R3] The Business rules associated with the creation and maintenance of these objects are defined in the ICS Collections Manual.

The following table identifies and describes the Object Modeling Technique(OMT) Rumbaugh notation that is used in the data models contained in Figures 4.7 - 4.15. Column one contains the OMT term, column two the graphical interpretation of the term and column three a brief description of the notation.

OMT Notation

Class-&-Object		Defines a class and its associated objects which provide an abstraction of something within the problem domain. The class contains attributes (data) and methods (functions) which are common between all objects in the class. Example: Collection is a class.
Generalization; **Specialization**		Defines generalization - specialization relations between classes. The specialization classes inherits the attributes and services of the generalization class. A generalization class can have one or many specializationÕs. A specialization can have one or more generalizations. Example: a Map Projection is a specialization of Planar System Class.
Whole-Part		Defines whole-part relations between classes. Example: Processing Options, Processing, and Scene Selection, are a part of Product Delivery Options.
Association **Name**		Identifies the relationships among classes... A relation has two labels, forward and inverse. For Example, Class 1 describes Class 2 is a forward direction,; the opposite would be Class 2 is described by Class 1 which is the inverse. These labels help to define the role of two or more classes with respect to the relation.

Multiplicity of Associations	 Class A Class B +1 Class C 1, 4, 25 Class D Class E	Multiplicity specifies how many instances of one class may be related to a single instance of an associated class. For Example An instance of a collection may relate to one or more Spatial Coverage's. The examples on the left express the ways in which multiplicities may be specified: Class A=many(zero or more) Class B= optional(zero or one) Class C=one or more Class D=numerically specified Class E=exactly one
Qualified Association	 Class 1 Qualifier Association Name role-1 role-2 Class 2	Role names identify the participation of the objects in the overall association. For example The relationship between Collection and Contact states that each Collection must have at least one Contact who is an administrator. Qualifier is an attribute which serves to further qualify the association. For example the association between the Product class and Product Delivery Options class is qualified by the attribute Group ID

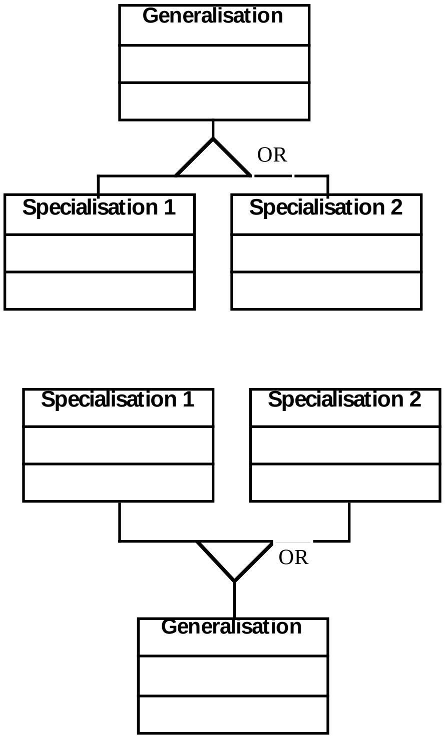
Exclusive Or	 Generalisation OR Specialisation 1 Specialisation 2 Specialisation 1 Specialisation 2 OR Generalisation	Exclusive "OR" indicates that only one of the specialization classes may exist for a given instance of the generalization class. This is graphically illustrated by indicating OR next to the specialization symbol.

Figure 4-7.1 reflects the collection module which illustrates the data objects, attributes, and associations related to the collection. This model reflects the breadth of classes and relationships that could exist for any instance of a collection. The off page classes, Spatial Reference, Spatial Coverage, Database Info(Explain) and Item Descriptors(Extended Services) are illustrated in other CDB figures, Explain Database or Extended Services Database figures. The Item Data and Browse Pointer are references to instances of browse information that are characterized outside of the CDB.

Figure 4-7.2 reflects the product module which illustrates the data objects, attributes, and associations related to the product. The same explanation provided above also applies to this module with the exception of the Database Info which does not relate to the product.

Figures 4-7.3 and 4-7.4 illustrates the spatial coverage and spatial reference modules respectively and their individual data objects, attribute, and associations. In the Spatial Coverage Module the Spatial Coverage class is a special purpose class which exists to add clarity to the module. It will not contain attributes and therefore will not have instances other than the indicated specialization's. The G Polygon is also reflected in this module to add clarity and will not contain instances other than through the specified relationships Outer G-Ring and Exclusion G-

Ring. The Spatial Reference Module contains several special purpose classes which are identical in purpose to the classes describe in the Spatial Coverage Module. The Horizontal Coordinate System, Vertical Coordinate System, Spatial Reference and Planar System are all special purpose classes which serve to add clarity/definition to the module.

Figure 4-7.5 reflects the Product Order information by illustrating the order objects and associated attributes. The relationship between this module and the Extended Services Database is also captured in the Order Module.

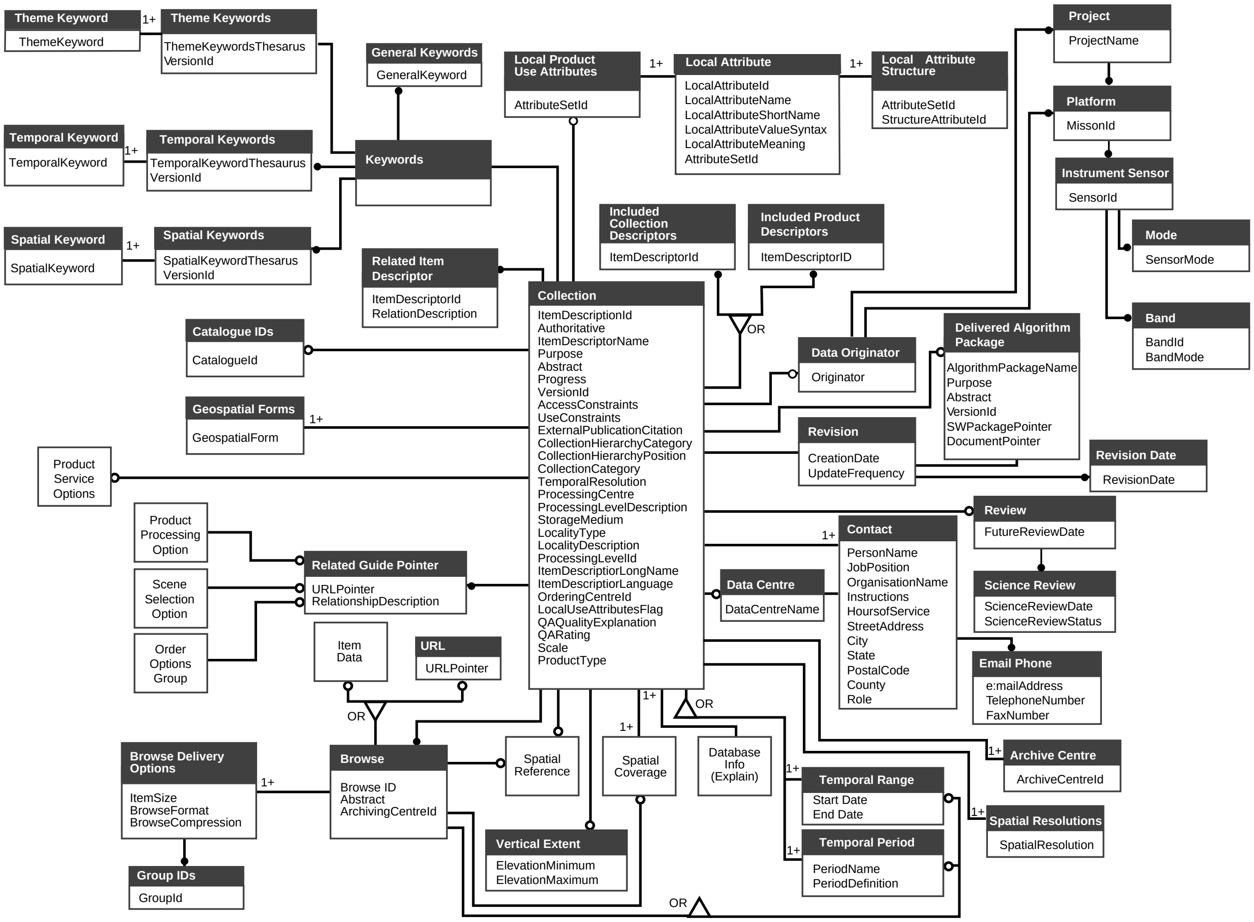


Figure 4-7.1. Collections Database Data Object Model - Collections Module

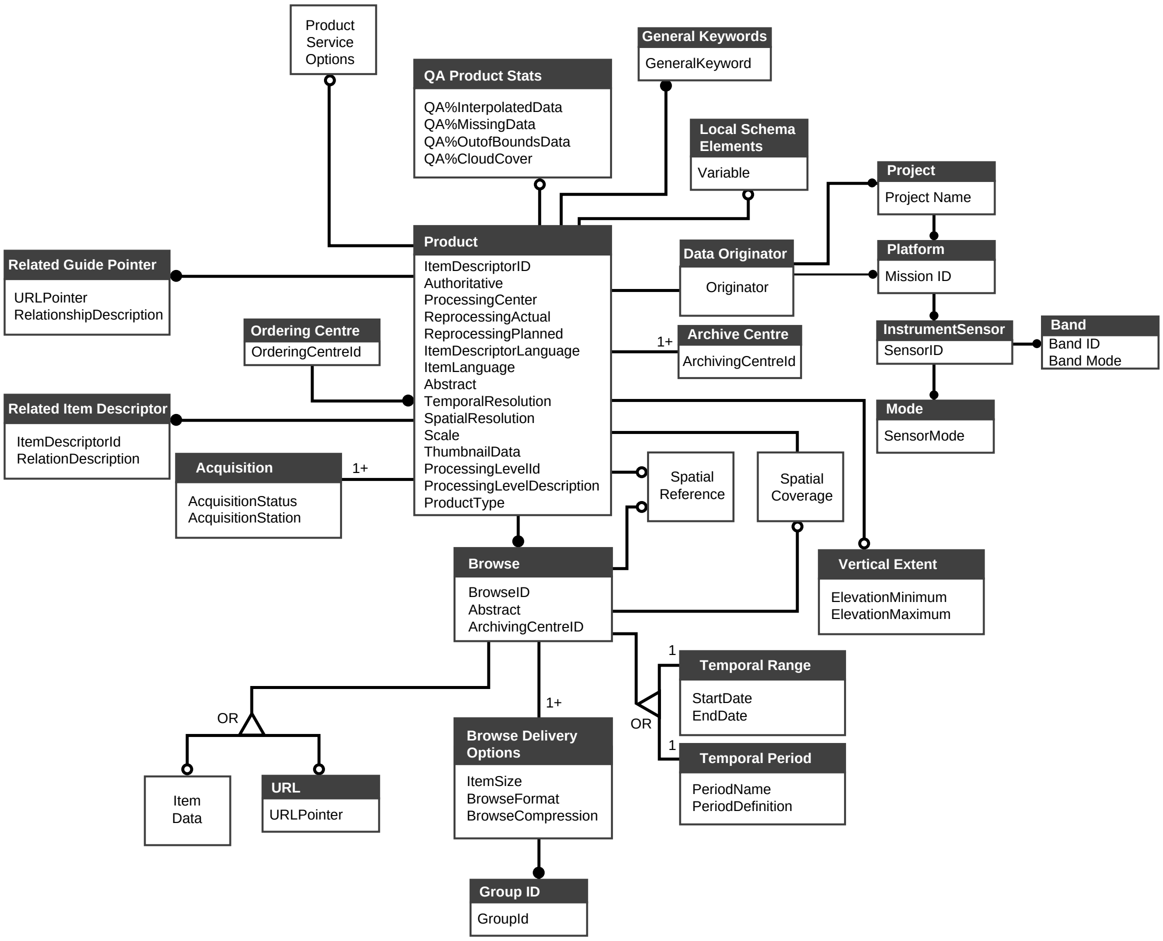


Figure 4-7.2. Collections Database Data Object Model - Product Module

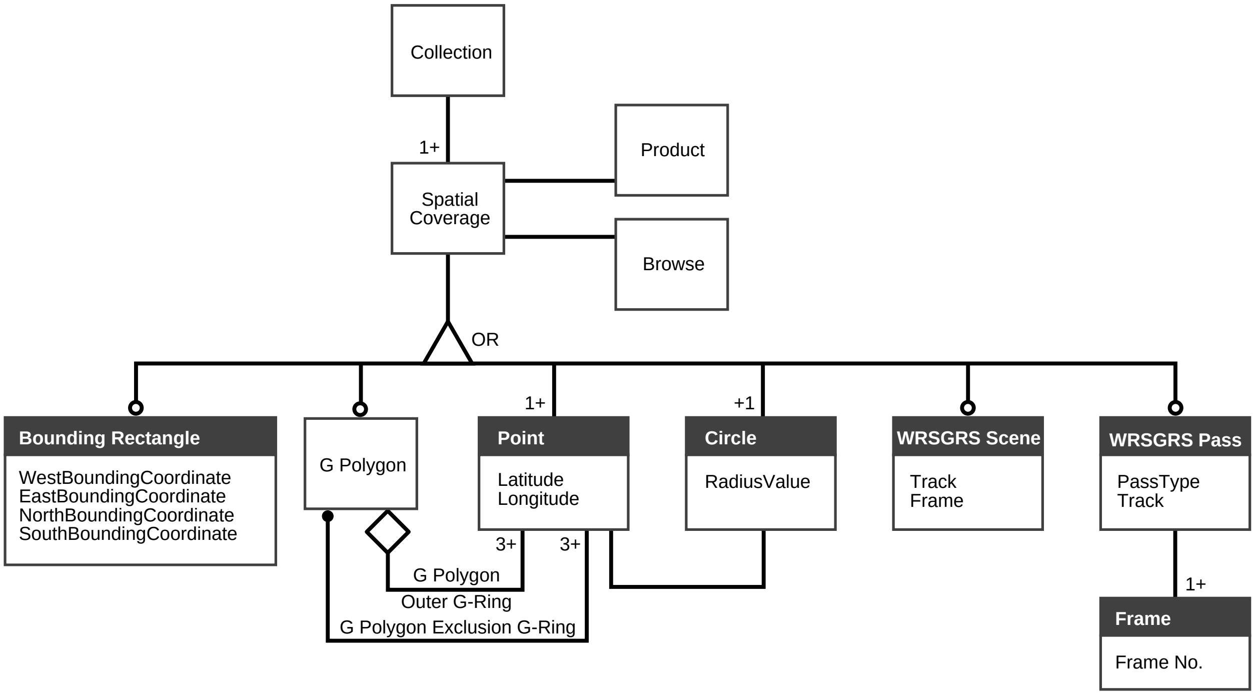


Figure 4-7.3. Collections Database Data Object Model - Spatial Coverage Module

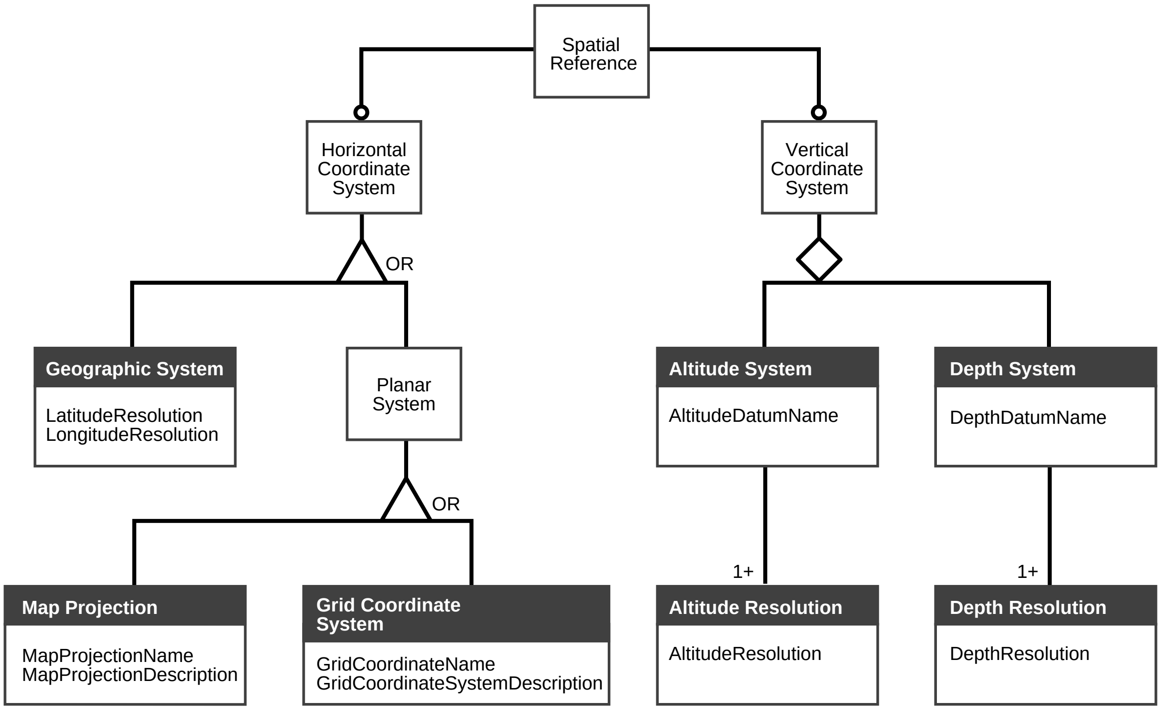


Figure 4-7.4. Collections Database Data Object Model - Spatial Reference Module

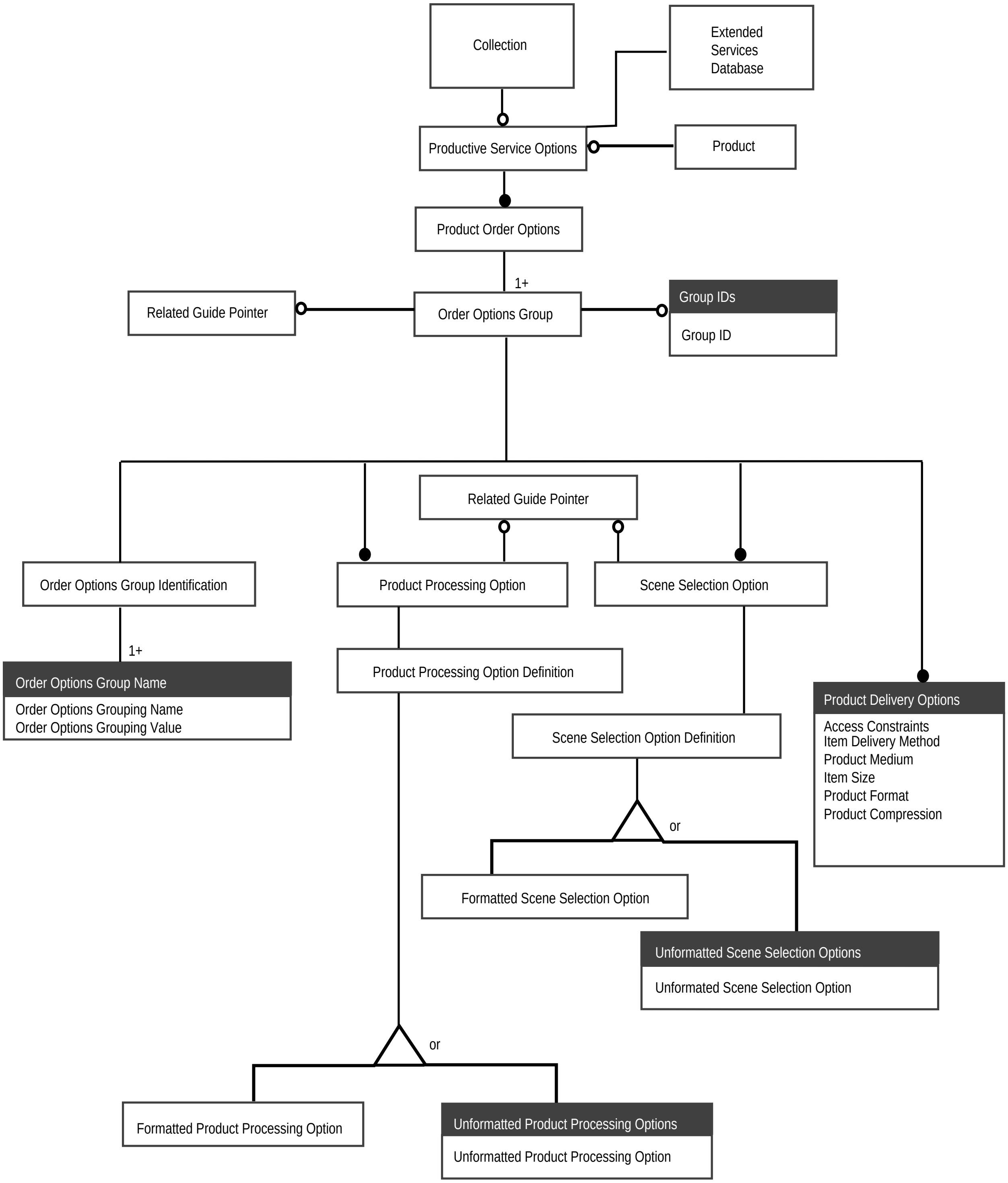

Collection
Extended Services Database
Productive Service Options
Product
Product Order Options
1+
Related Guide Pointer
Order Options Group
Group IDs
Group ID
Related Guide Pointer
Order Options Group Identification
Product Processing Option
Scene Selection Option
1+
Order Options Group Name
Order Options Grouping Name
Order Options Grouping Value
Product Processing Option Definition
Product Delivery Options
Access Constraints
Item Delivery Method
Product Medium
Item Size
Product Format
Product Compression
Scene Selection Option Definition
or
Formatted Scene Selection Option
Unformatted Scene Selection Options
Unformated Scene Selection Option
or
Formatted Product Processing Option
Unformatted Product Processing Options
Unformatted Product Processing Option

Figure 4-7.5. Collections Database Data Object Model - Order Module

4.6.1.6 Collection Census

ICS Compatability:Explanatory

This section provides statistics on the number of collections anticipated in ICS coupled with estimated metadata sizing per product.

It order to support the design analysis, a rough order of magnitude estimate of the number of collections in ICS is needed. A specific count of the collections is not necessary nor is a list of named collections. The estimates provided here are also not aimed at indicating an expectation that a particular agency will provide access to the number of collections indicated. The purpose of the estimate is to provide a high end estimate of what the CEOS collection space might grow to. This estimate is provided in Table 4-1.

The estimate reflected in Table 4-1 was discussed during the CEOS Catalogue Sub-Group meeting in May 1996. (The Catalogue Sub-group has been renamed the Access Sub-group.) It was the conclusion of the Access Sub-Group that the estimate in Table 4-1 significantly under estimated the number of collections which could be indexed using CIP. In particular, the GCMD experience indicates that there are millions of datasets, although a large portion of these datasets will never be accessible on-line. The number of datasets which may provide on-line directory metadata may be on the order of hundreds of thousands. And for datasets which would have inventory on-line, an estimate on the order of tens of thousands was felt to be accurate.

For now the best that can be said is that CIP should be sized for hundreds of thousands of collections with collection level metadata only. Further, CIP should be sized to hold tens of thousands of Collections containing product descriptors.

Table 4-1. ICS Collection Upper Bound for Design Sizing

Agency	Rough Estimate of the Number of Archive collections
European	
BNSC	**100 +/- 50**
CEO	**50 to 500**
CNES	
DLR	
ESA	**50**
Canadian	
CCRS	**500**
Japanese	
NASDA	**70**
US Agencies	
NASA (EOSDIS)	**Version 0:1000 Version 1: 2000**
NOAA	**1200**
USGS	
Total	**Roughly 5,420**

Also important to ICS element sizing is the size of product metadata, i.e. the size of a product descriptor. This number will certainly vary for the various products. For a first order estimate, an estimate from ECS is provided here. ECS is using an estimate of 2K/product. This estimate is being used for both ECS Release A (predominately Version 0 and TRMM data) and for the ECS DBMS prototype for ECS Release B (Release A plus EOS AM-1, Landsat-7, and others).

4.6.2 Explain Database

The Explain Database offers the ICS the structure and information necessary to respond to the Z39.50 Search, and Present Services. This component of the ICS data architecture “Explains” the data or information environment in which the ICS will operate. This environment consists of categories of information that the target supports. These categories would include databases, schema’s, record syntax, attribute sets, and extended services. For each of these categories, a detailed description, which is contained in the attributes, relationships, and associated data objects serve to provide the model under which the ICS will capture, store, and make available information about the ICS *Retrieval Manager*. The Explain database data model which is illustrated in Figures 4-11.1 to 4-11.2 graphically depicts these concepts in addition to the relationship between Explain and the Collections Database.. Additional details regarding each of the objects in the data model are described in the CIP Specification - Release B [R3]. The creation and maintenance of the Explain entries are described in the ICS Collections Manual.

4.6.2.1 Explain Database Data Model

ICS Compatability: Mandatory

The following briefly describes each of the data objects referenced in the Explain Data Model Figure 4.11.1. Collectively these objects serve to organize the Explain data into manageable entities which can be readily understood by data engineers.

Database Info
This object describes each collection in the logical CDB. The ISA must create an Explain Record in the Database Info for each CDB collections entry. This explain record should include the database name, which must be described according to the specification in the CIP Release B Specification [R3], Appendix D.2, Database Names. Access available, Record Count, Average Record Size, User Fee, and Max Record Size should also be specified in this record. Additionally, if the collection is a Root Collection then the content of the Keyword Attribute should reflect Root.

Schema Info

The Schema Info Data Object identifies the abstract record structures which are the elements that describe the collection referenced in the Database Info. There are currently two (2) record structures that are supported in the ICS:

- Collections
- Products

Each Database Info entry must relate to at least one of the above entries in the Schema Info.

Retrieval Record Details

This object contains the descriptive details about the elements of the retrieval record.

Record Syntax Info

The Record Syntax Info is used by the client to determine the format of the retrieval record. Four formats are currently available; Explain; Sutrs; GRS-1; and Extended Services. Each entry in the Database Info must relate to one or more Record Syntax's. The Record Syntax Info coupled with the Element Set Details (described below) allows the client to determine both the format and content of the retrieval record.

Element Set Details

This data object allows the client to determine the data that will be retrieved by specifying the Element Set Name. To achieve this the Database Info and Schema Info Records which were created for each collection must be related to an instance of the Element Set Details. Each Element Set Details is then related to an instance of an element(s) in the Per Element Set Details, Explain Records. For example the Element Set Detail Name = "FULL" has been related to the Per Element Set Details; Element Name = Item Descriptor ID for the Schema; Name ="Collection" for the Database Info;Name="XYZ"

There are six element set names:-

- Full-
- CIP Full-
- Brief-
- Browse-
- Collection Member-
- Options-
- Local Attributes-
- Summary-

Per Element Set Details Details
This data object contains a description of each element and is related to the instance of the Database Info through the Schema Info and Element Set Details.

Attribute Set Info
The Attribute Set Info provides a description of each attribute set supported by the target. This set includes the OID, Name and description of the attribute set.

Attribute Type
This specifies the type of attribute i.e. Use Attribute, Truncation Attribute, Relation Attribute, etc. This object also includes the name and description of each attribute type.

AttributeDescription
Specifies the description of the attributes associated with the attribute type. For example an attribute type = “Relation” may have an attribute description = “Less Than” and Attribute Value = “1”. The attribute value is the ID# associated with the attribute description.

Query Types
This object specifies the query types supported by the RM. The RPN is the only Query Type currently available in the CIP.

Unit Info
The Unit Info specifies the unit system supported by the target system. The ISA establishes this information from the contents of the data that the RM is supporting.

4.6.2.2 Local Attribute Definition Using Explain

ICS Compatability: MAA

The Explain Database may be used to capture the definitions of the local attributes rather than the Collections Database (ref Section 4.6.1). The process for recording this information in Explain is described in detail in the Collections Manual [R5]. The important concept within the SDD is that the Explain Database will not require additional attributes, objects or relationships beyond what is specified in the Explain Data Model Figure 4.11.1, to support this requirement.

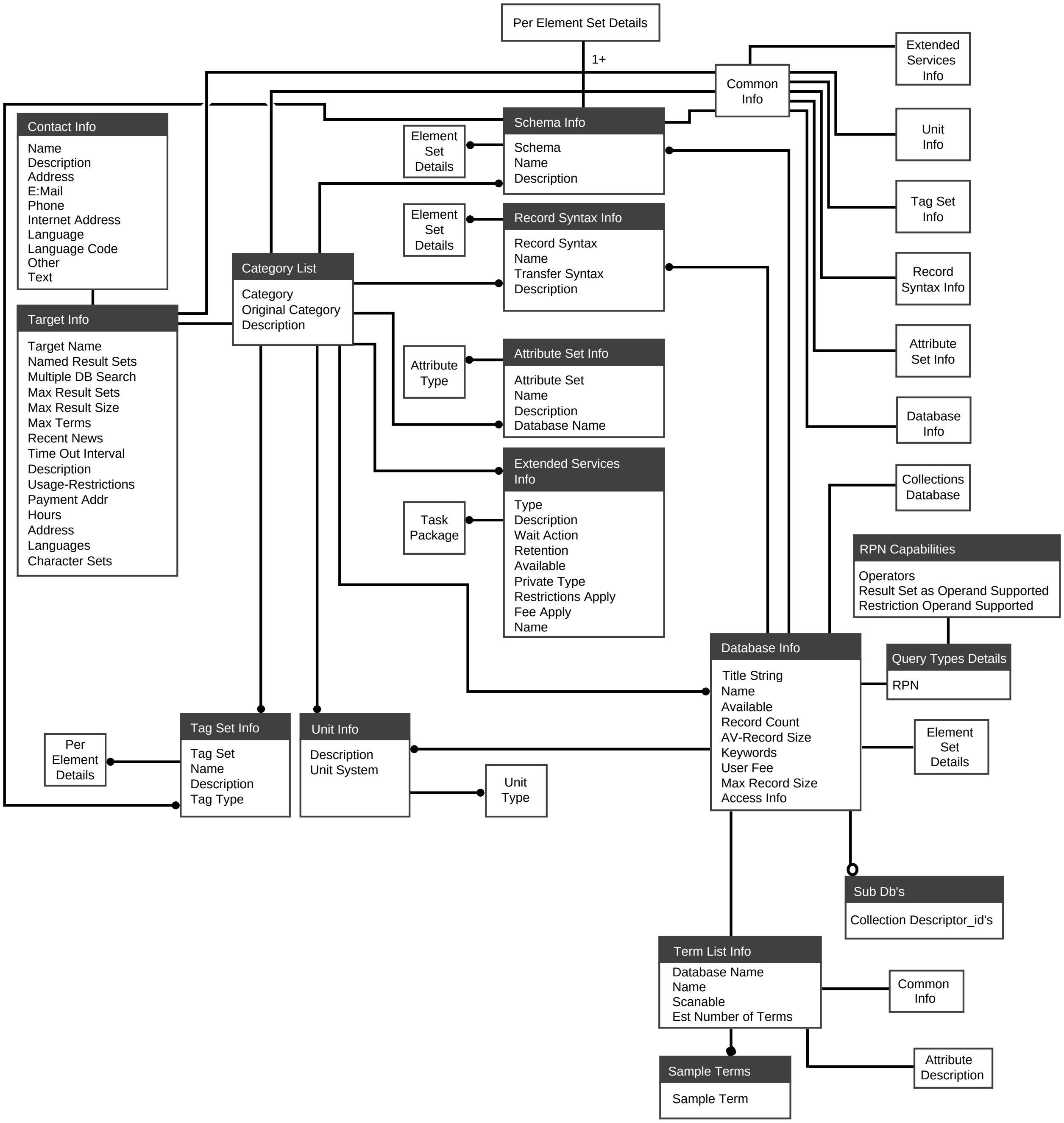


Figure 4-11.1. Explain Database

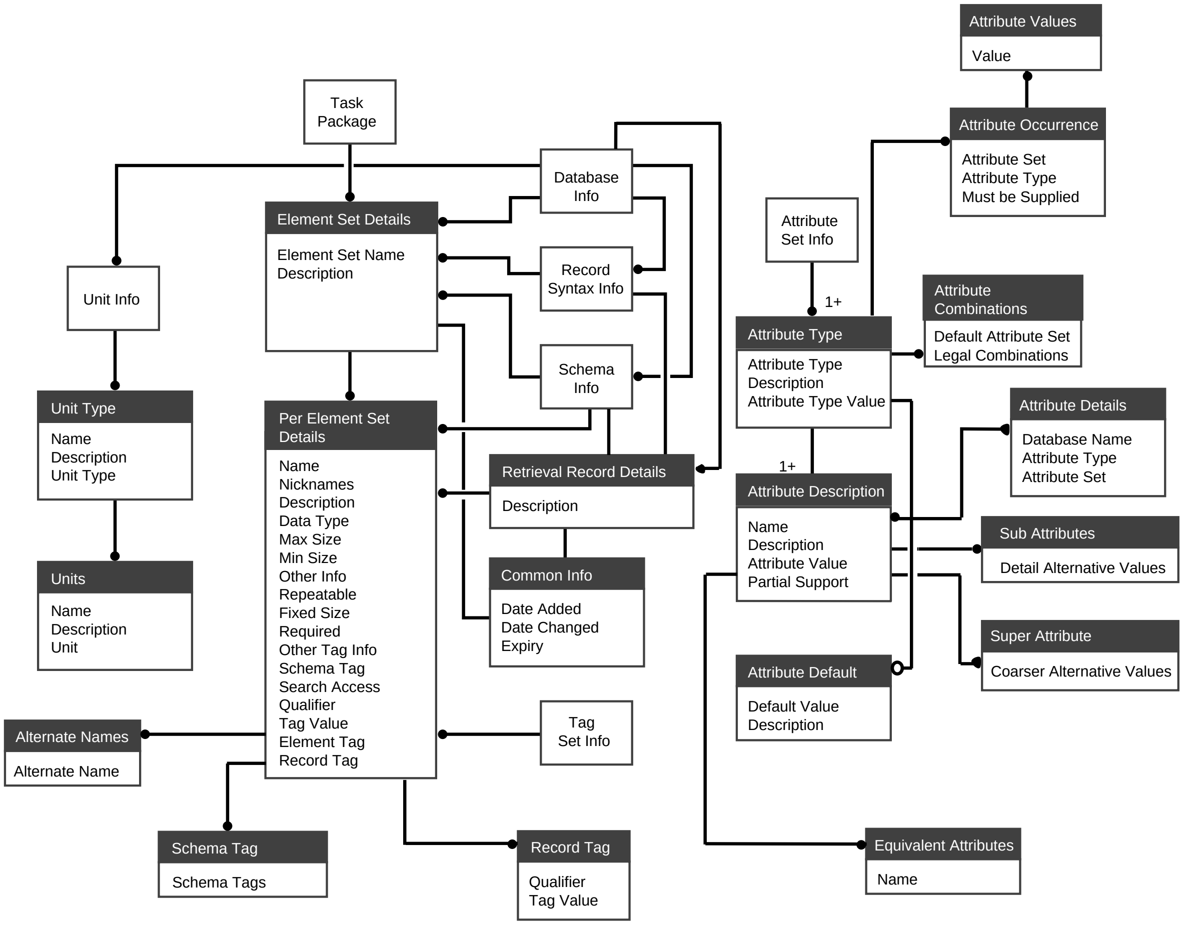


Figure 4-11.2. Explain Database

4.6.3 Extended Services Database

ICS Compatability: MAA

The Extended Services Database component of the ICS Data Architecture provides the information necessary to support the Z39.50 Extended Services Facility. This facility offers to ICS the extensions referred to as Task Packages, necessary to support the persistent storing and accessing of Order, Query, Database Update, and Result Set information. The Extended Services Database contains a description of each of these task packages. The persistent result set task package contains the administrative information about persistent result sets This would include the package ID which would be the ID of the Theme Collection referenced in the collections database.The Order Task Package contains the information necessary to respond to an order request from an origin. The Query Task Package provides the metadata associated with the persistent queries which are stored in a query file. Lastly, the Database Update Task Package provides a journal of the database activity associated with an ICS database. Each of these packages may have assigned a User or a Group of Users and their assigned privileges. Figure 4-12 graphically illustrates the Extended Services Database data object model. Additional details regarding each of the objects in the data model are described in the CIP Specification - Release B [R3]. Beyond these details the relationships between the extended services and the collections are illustrated in Figure 4-12.

The following briefly describes each of the major objects referenced in Figure 4-12.

TaskPackage

This specifies the administrative information (status, time of creation etc.) of the task package in the Extended Services Database.

Result Set

This object contains the characteristics of a persistent result set to include the Result Set Name and the Number of Records contained in the result set.

OrderSpecification

The order specifies the characteristics about a specific order such as the status, order date, Delivery Method, Package Medium, Pricing Information, and Billing Method.

Persistent Query

This object contains the characteristics associated with a persistent query such as the definition of the query, parameters associated with the execution of the query. This object is related to the database info object as the persistent query must target a collection.

Periodic Query Schedule

This object specifies the characteristics associated with a query schedule. This query schedule relates to both the Persistent query that it may execute and persistent result sets which the persistent query may generate.

DB Update

Specifies database update activity for a specified database.

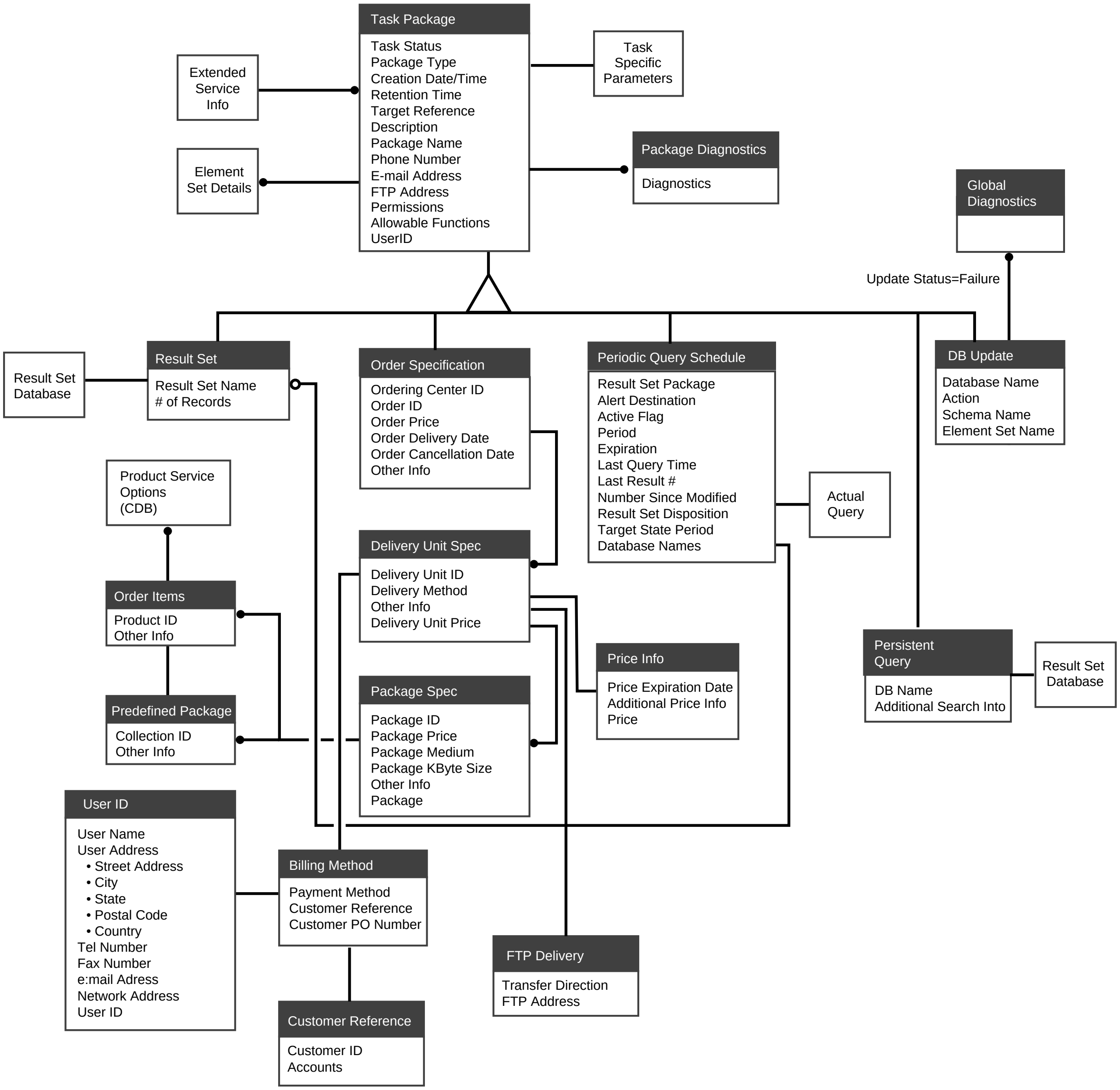


Figure 4-12. Extended Services Database

4.6.4 Persistent Result Set Database

ICS Compatability: MAA

The Persistent Result Set DB contains references to either collections that are in the result set or queries, that when executed, dynamically generates the list of collections. These references may be local or remote or a combination of both. When a collection is encountered which includes a task package reference by including an IncludedCollectionDescriptors, which references a persistent query, an extended services request is generated that targets the extended services database using the packageName referenced in the collection. An extended services response is generated in response to this request which contains a RPN Query. The RM appends the RPN Query(persistent query) to the original query and executes. The result set contains the intersection data, that is the data that satisfies both the persistent query and the original query.

The following Result Set Database Data Model captures the relationships between the Extended ServicesEnvironment, Result Set Database, Local Inventory and the Collections Database.

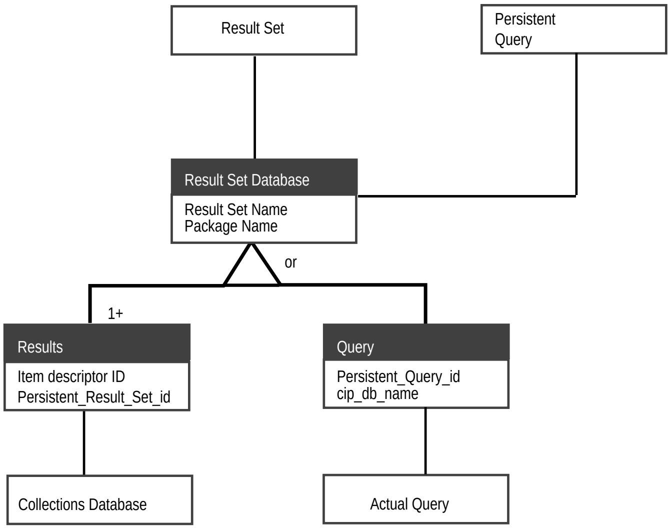


Figure 4-13. ResultSet Database Data Model

4.6.5 Session Management Database

ICS Compatability: Mandatory

The session management data will contain the information necessary to support Query, Results and General Session activity. The intent of this data are to provide the necessary information to support the interaction between the *Retrieval Manager* and the interfaces external to the *Retrieval Manager*.

- Query Data is characterized as information required to support the query processes to include both search and present. Additionally, Query Management information would also be included which would reflect Routing, and Status of Queries
- Results Data will include all of the information required to support results processing.
- General Session Information will include User Interaction Session logs which would contain the User ID, Location, Start and End Times, Profile of Tasks Performed, and Resources Used. Additionally, the CIP User Log which would capture Authentication Attempts, Successes, Failures, Service Options Accessed and Unsuccessful Log Ins would also be included, as well as command control information such as types of requests (order, batch, status query).

4.6.6 Error Management Database

ICS Compatability:Explanatory

The error management data are the data required to provide a description of error conditions which may occur during processing. This includes diagnostic messages, level of priority of error or failure and the location of the error within the system to include attribute error, software error, or hardware error. CIP Diagnostic Messages are listed in Appendix E of the CIP Specification - Release B [R3].

4.6.7 User Management Database

ICS Compatability:Explanatory

User Management data includes the User Profile information including the data required for authentication and user privileges. Additionally, *Retrieval Manager* profile data such as locations are also included in the user management database. The following model represents the user informatiion that will be captured in the User Management Database.

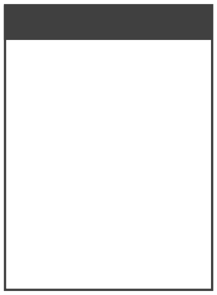


Figure 4-14. User Descriptor Model

4.7 IGP Document Data (Guide, Reference Papers etc.)

Although the Document protocol supports the insertion and retrieval of any type of document, the most common example from the Earth Observation (EO) perspective is the Guide Document. Guides are textual documents that are developed to assist the user in understanding the EO environment. Guide Documents provide additional detailed information about a specific mission, sensor, or organization. The most frequently covered guide topics include:

Data Centers - provides details about the center that holds the data.

Project/Campaign - describes projects and associated intensive field campaigns coordinated to collect data for a focused study.

Source/Platform - describes the source that held the instrument and sensors during data collection. Source is intended to communicate the data collection environment which includes satellite, aircraft, buoys, ships and ground station platforms. It also includes humans in the case of hand held sensors or human observations, paper or electronic surveys in the case of questionnaires, and computers in model analysis.

Sensor/Instrument - describes the instrument and its component sensor(s) that actually collected the data. This includes eye as a sensor, as well as, paper or electronic questionnaires used to collect data. It also includes computer model analysis that generate data.

Dataset - describes the dataset, collection, procedures, algorithms, and processing data.

Document Data will be captured in the local systems environment. The details for accessing and retrieving this information is specified in the ICS Guide Design and Protocol Specification [R24].

4.7.1 Guide Document Data Model

ICS Compatability: MAA

The following data model reflects the guide data for the Guide Indexer referenced in the ICS Guide Design and Protocol Specification [R24]. This model identifies each of the objects and further attributes that will be supported in addition to the relationships between the data objects.

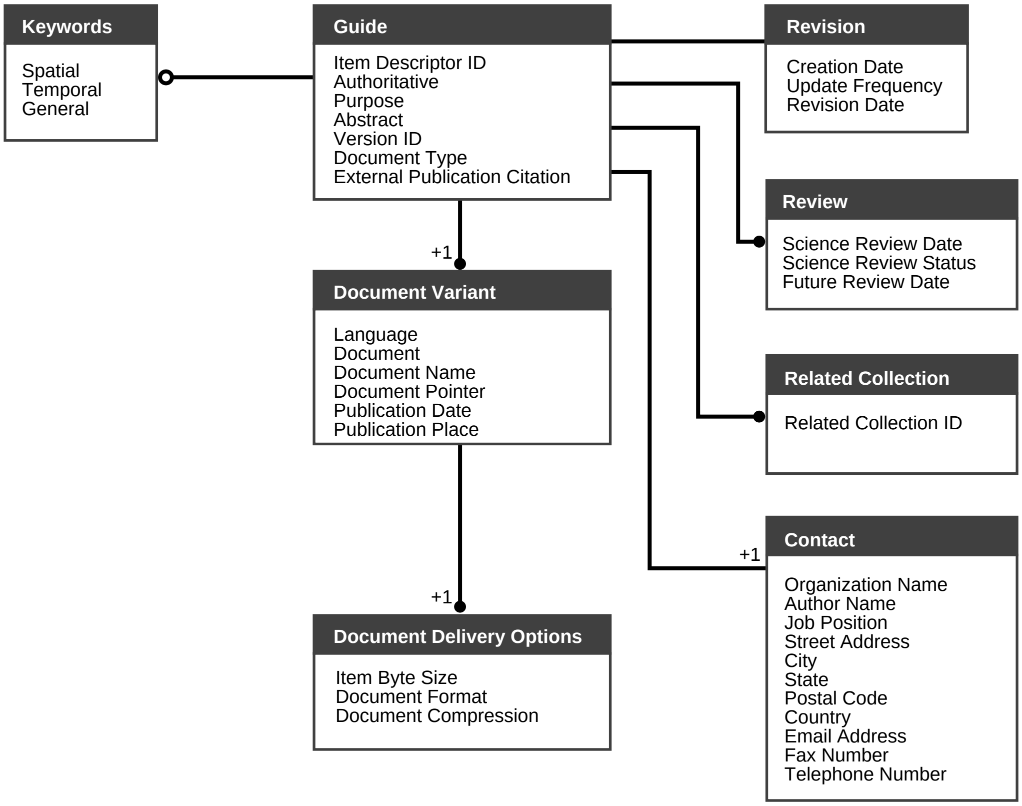


Figure 4-15. Guide Data Model

This page intentionally left blank.